Morphosyntactic Processing in Late Second-Language Learners

Margaret Gillon Dowens[1,2], Marta Vergara[3], Horacio A. Barber[1], and Manuel Carreiras[4,5]

Abstract

■ The goal of the present study was to investigate the electrophysiological correlates of second-language (L2) morphosyntactic processing in highly proficient late learners of an L2 with long exposure to the L2 environment. ERPs were collected from 22 English–Spanish late learners while they read sentences in which morphosyntactic features of the L2 present or not present in the first language (number and gender agreement, respectively) were manipulated at two different sentence positions—within and across phrases. The results for a control group of age-matched native-speaker Spanish participants included an ERP pattern of LAN-type early negativity followed by P600 effect in response to both agreement violations and for both sentence positions. The late L2 learner results included a similar pattern, consisting of early negativity followed by P600, in the first sentence position (within-phrase agreement violations) but only P600 effects in the second sentence position (across-phrase agreement violation), as well as significant amplitude and onset latency differences between the gender and the number violation effects in both sentence positions. These results reveal that highly proficient learners can show electrophysiological correlates during L2 processing that are qualitatively similar to those of native speakers, but the results also indicate the contribution of factors such as age of acquisition and transfer processes from first language to L2. ■

INTRODUCTION

In a world increasingly more pluralistic, multicultural, and multilingual, the study of second-language (L2) acquisition in adulthood is a research area that has become increasingly important not only in its own right but also due to the significant social and educational implications of the findings from these studies. However, despite the recent increase in research in this area, little is still known about how an L2 acquired after childhood is represented, how it is processed, and whether this is indeed qualitatively different to what happens in childhood L2 acquisition. There is a common assumption that L2 attainment is primarily or even exclusively a function of age and that native-like performance is not attainable unless the L2 is acquired during early childhood (Eubank & Gregg, 1999; Johnson & Newport, 1991). This idea of a maturationally constrained critical or sensitive period for L2 acquisition underlies, for example, the failed functional features hypothesis (Hawkins & Chan, 1997) and the shallow structure hypothesis (Clahsen & Felser, 2006b) among other models of L2 acquisition. However, recent experimental data do not unequivocally bear out this observation (Rossi, Gugler, Friederici, & Hahne, 2006; for an overview, also see Birdsong, 2006), and there is now a growing awareness that there is not one but a number of different factors that can influence the learning of an L2 after childhood and the levels of competence and performance attained in this language. These factors may include language proficiency (Ojima, Nakata, & Kakigi, 2005), similarity of the syntactic structures of the first language (L1) and L2 (Jeong et al., 2007), and limitations of general cognitive resources such as working memory (McDonald, 2006). There is, thus, a need for more detailed information and more empirical evidence about the nature of these different factors affecting adult L2 acquisition, how they interact, and how they apply to different language components. To answer these questions with precision, there is a particular need for on-line and neuroimaging studies of late L2 processing to complement the behavioral research carried out so far. This is the aim of the present ERP study.

Since Lenneberg (1967) and Penfield and Roberts (1959) first hypothesized the existence of a biologically circumscribed period for L2 acquisition, there have been many reformulations of this critical period hypothesis (CPH; for overviews, see Birdsong, 2006; DeKeyser & Larsen-Hall, 2005; Hyltenstam & Abrahamsson, 2003), but in most of them, the basic tenet is still that the ability to acquire and to produce the L2 in a native-like manner is maturationally constrained begins to decline sometime in childhood and has disappeared by puberty. However, there is intense and continuous debate about the nature, the specifics, and indeed the validity of this concept

[1]University of La Laguna, Spain, [2]University of Nottingham Ningbo, China, [3]University of California, Davis, [4]Basque Research Centre for Cognition, Brain and Language, Spain, [5]UVP/EHU

(Singleton, 2005; Birdsong, 1999). At present, there are two main arguments against the CPH. First, it is open to question whether there is indeed a critical "period" or developmental window as such for L2 acquisition and not just a general decline across the age span in the ability to learn some aspects of language, mirroring the age-related decline in other cognitive resources such as working memory, some which are involved in language processing (Bialystok & Hakuta, 1999). Second, there is growing evidence of late L2 learners with native-like performance in on-line as well as off-line tasks, which casts doubt on the basic premise of the "impossibility" of L2 acquisition beyond the offset of any proposed critical period (van Boxtel, 2005; Marinova-Todd, 2003). Thus, the idea of a general critical or sensitive period for L2 learning is now often replaced by arguments that there are some particular areas of language processing, such as phonology and syntax, which can never become native-like if they are acquired after childhood or adolescence. L2 attainment in the area of syntax, particularly, is considered to be constrained by age of acquisition (AoA) more than other areas such as lexical processing, and so it is argued that in this area, only early (childhood) bilinguals will achieve native-like levels of proficiency and show on-line processing patterns similar to natives (Clahsen & Felser, 2006a, 2006b; Weber-Fox & Neville, 1996). The data, however, are far from consistent, with other, more recent studies painting quite a different picture (Rossi et al., 2006; Mueller, Hahne, Fujii, & Friederici, 2005; Ojima et al., 2005). That there is conflicting evidence is hardly surprising, given the nonhomogenous nature of the population under study, the different tasks, and the experimental techniques used in L2 acquisition research and indeed the lack of general agreement about the characteristics of bilingualism and the evaluation of bilingual language processing (Grosjean, 1989). Also, compared with the equivalent L1 research, there have been relatively few on-line studies of late L2 morphosyntax processing to date, so more information is needed to clarify this question of the relative importance of AoA. What is increasingly clear, however, is that age is not the only factor to be taken into consideration. One other major factor that has received much attention recently is the question of L2 proficiency. Many earlier on-line studies reporting non-native-like processing patterns either did not control for proficiency (Weber-Fox & Neville, 1996) or studied late learners with relatively low exposure/low proficiency in the L2 (Hahne, 2001). When proficiency is controlled for, the results can be quite different (Rossi et al., 2006; Abutalebi, Cappa, & Perani, 2001). Another factor that has received attention recently is the question of L1–L2 language similarity and feature overlap, as different results have recently been obtained from studies using different L1–L2 language pairings (Sabourin, Stowe, & de Haan, 2006). Also to be taken into consideration is the role of other more general cognitive components and resources such as memory resources in L2 language processing (McDonald, 2006).

So far, the relative importance and influence of these different factors is largely unresolved. Our intention is to cast some light on the influence of AoA, L2 proficiency, and language feature overlap by looking at late learners' morphosyntactic processing of L2 sentences, specifically processing of gender and number agreement features. We registered the on-line brain responses to violations of these agreement features of a group of highly proficient late (postpuberty) learners of Spanish whose L1 is English and who have acquired the L2 by immersion in the L2 environment while living in Spain for over 20 years, on average. The off-line tests and self-rating questionnaires of these late learners indicate high L2 competence, but on-line, do they process the L2 like natives?

We registered the participants' responses to these grammatical violations using the ERP technique. The excellent temporal resolution of this technique makes it particularly suitable for the study of how and when agreement relationships are computed during reading. Of particular interest are two ERP components that are commonly found in studies of morphosyntactic processing. The first of these is a LAN, a negative deflection in the brain wave occurring between 300 and 500 msec after stimulus onset, usually with a more anterior scalp distribution and a left hemisphere bias (Friederici, Hahne, & Mecklinger, 1996; Kluender & Kutas, 1993), although other distributions have been commonly reported (Silva-Pereyra & Carreiras, 2007; Hagoort, Wassenaar, & Brown, 2003; Kluender & Münte, 1998; Münte, Matzke, & Johannes, 1997). LAN-type effects have been reported for different syntactic violations, including number and gender mismatches (Barber & Carreiras, 2005). It has been argued that the LAN effect could be an index of the detection of a mismatch between morphosyntactic features (Rodriguez-Fornells, Clahsen, Lleó, Zaake, & Münte, 2001; Gross, Say, Kleingers, Clahsen, & Münte, 1998), of the difficulty of integrating these characteristics in a syntactic structure (Hagoort, 2003; Friederici, Steinhauer, & Frisch, 1999), or of increases in the working memory demands implied in these processes (Vos, Gunter, Kolk, & Mulder, 2001; King & Kutas, 1995; Kluender & Kutas, 1993).

The second component of interest is the P600, or syntactic positive shift, a widely distributed positive deflection with posterior maximum, an onset at around 500 msec after stimulus presentation and a long duration, often around 400 msec. The P600 has been widely reported in different languages for a range of syntactic violations, including number and gender mismatches (Hagoort, 2003; Barber & Carreiras, 2005). As to the significance of this effect or family of effects, as some researchers have suggested might be the case, the P600 was originally considered to be related to syntactic processing, reflecting different stages of the detection, reanalysis, and repair of syntactic anomalies (Friederici, Hahne, & Saddy, 2002; Hagoort & Brown, 2000) or processing of syntactic complexity. However, there has also been recent evidence associating this component with semantic/thematic processes (van Herten, Kolk, & Chwilla, 2005) and so some researchers now believe that this component may reflect less domain-specific processes associated to anomaly detection or general monitoring processes (Coulson, King, & Kutas, 1998). One explanation of the LAN–P600 ERP pattern in response to syntactic violation is that proposed by Friederici et al. (2002). In this proposal, an early LAN component between 150 and 200 msec indexes an initial, automatic stage of phrase-structure building based on information about word category, a later LAN between 300 and 500 msec indexes a stage of morphosyntactic integration, whereas the P600 indicates a later, more strategic stage of reanalysis and repair of the structure (on the timing of these syntax-related ERP effects, see, however, Hasting & Kotz, 2008; Hasting, Kotz, & Friederici, 2007). The LAN–P600 pattern is, then, a well-documented ERP experimental result for native speakers processing morphosyntactic agreement errors, but what about late L2 learners, our group of interest? LANs have rarely been observed in studies of late L2 learners, which has often been taken as evidence for the absence of native-like processing and as support for critical period constraints. However, LAN effects have been observed in some artificial L2 grammar studies (Mueller et al., 2005; Friederici, Steinhauer, & Pfeifer, 2002), which would suggest that some native-like syntactic L2 processes can be acquired by adults after intense training, and at least one previous study has reported a LAN effect for highly proficient late learners in a natural language (Rossi et al., 2006). The results of the few on-line studies of L2 gender and number agreement processing to date have been somewhat contradictory, possibly due to the abovementioned reasons. Sabourin et al. (2006), in an ERP study on gender agreement in L2 Dutch, compared L1 English, L1 German, and L1 Romance adult learners. Only the L1 German group evinced a native-like P600 effect; the L1 Romance group showed a late frontal negativity and the L1 English group showed no robust effects at all. However, in another recent ERP experiment on gender agreement in L2 French (Frenck-Mestre, Osterhout, McLaughlin, & Foucart, 2008), the authors found that L1 English speakers of French do evidence ERP signatures of gender agreement violations, although these ERP responses to agreement violations are not fully native-like.

To study the question of L1 influence, feature overlap, and transfer, we will look at two types of morphosyntactic agreement features of the Spanish language—number and grammatical gender—as the former exists in the L1 (English) and the latter does not. Although number is computed less in English than in richly inflected languages such as Spanish, there is number agreement in English, for example, between determiners and nouns ("that dog/those dogs"). However, there is no equivalent in English to the grammatical gender assignment of Spanish, where each noun is either masculine or feminine and requires the agreement of all modifiers. As far as we are aware, there have been very few on-line experimental studies of L1 transfer effects (for an overview, see Kotz, 2008). Tokowicz and MacWhinney (2005), in a study of low-proficiency late L2 learners, found P600 effects for an L1 L2 similar syntactic construction and for an L2 unique construction but no effects where the syntactic feature was present in both L1 and L2, but the construction was differently formed. They reported this as evidence for L1–L2-positive and -negative transfer effects at least at early stages of L2 acquisition. Similarly, Sabourin and Stowe (2008) also reported that rule-governed constructions that are similar in L1 result in "native-like" P600 effects but not constructions that are grammatically different in L1 and L2. Again, this would indicate positive transfer effects for shared features and negative transfer effects for features that are not the same in the L1 and L2. Thus, if we obtain differential effects for the gender and the number feature manipulations, this could shed more light on this issue of L1–L2 feature similarity and transfer. Our manipulations should also provide evidence for or against one of two contrasting theories of L2 acquisition. The first of these, the failed functional features hypothesis (Hawkins & Chan, 1997), in concurrence with the CPH for L2 acquisition, posits that in late L2 acquisition only those features previously present in the L1 will be transferred or mapped onto the L2. Grammatical features not originally present in the L1, however, will not be acquired by L2 learners. In contrast, the full transfer/full access hypothesis (Schwartz & Sprouse, 1996) argues that although the L1 grammar is the basis for L2 acquisition, with the continual exposure to L2 and the achievement of high levels of proficiency, even grammatical features and processing routines not originally present in the L1 can become functional in the L2. We will be able to test these contrasting theories by examining the ERP response to violations of the L2 grammatical agreement feature already present in the L1 (number) and contrasting this with violations of the L2 grammatical agreement feature not present in the L1 (grammatical gender).

Some researchers believe that more general cognitive factors such as demands on working memory could be an important factor underlying non-native-like performance in late learners. In a previous study of gender and number agreement carried out with Spanish native speakers (Barber & Carreiras, 2005), the authors manipulated agreement violations in two sentence positions, with a within-phrase violation between article and noun at the beginning of the sentence and with a between-phrase violation between a noun and its postmodifying, postverbal adjective in the middle of the sentence. They found no differences between these sentence position violations for native Spanish readers. We maintained these two sentence position manipulations to see whether the added demands on working memory in the later sentence position would be more problematic for the late learners than for the native speakers. For example, in the sentence "El suelo está plano y bien acabado" (The floor is flat and well finished), the agreement relationship is manipulated between the article and the noun "El-suelo" (first position/within-phrase manipulation) or between the noun and the adjective "suelo-plano" (second position/across phrase manipulation). Presumably, the latter manipulation could be more costly as it requires maintaining the agreement features in memory for longer and across the phrase boundary. This has not proved to be problematic for native speakers, but the higher cost of processing in the L2 in terms of working memory is a well-documented phenomenon (McDonald, 2006).

Based on these previous results from studies of gender and number processing in L1 Spanish (Barber & Carreiras, 2005) and predictions from L2 acquisition models, we could expect some of the following results from our experiment: First, if AoA is the most relevant factor, as indicated by models based on the CPH, then we should not expect a native-like ERP pattern from these postpuberty L2 learners. There should be no LAN-type effects indicating native-like processing of the grammatical features (for a discussion of the timing of LAN effects, see Hasting & Kotz, 2008) and there may or may not be later P600 effects. If, on the other hand, proficiency is the key factor, then a more native-like pattern of LAN followed by P600 effects could be expected from these high-proficiency late learners. Second, if, as indicated by the failed functional features hypothesis, only the grammatical features are already present in the L1 map over onto the L2, then as number is computed in English but grammatical gender is not, any significant effects should only be for the number agreement violations. However, if, as argued by the full transfer/full access hypothesis, the features of the L1 are the starting point but not the limit of L2 acquisition, then after such prolonged exposure to the L2 and the high levels of proficiency obtained, there may be effects of both gender and number violations. There could also be differential effects for the gender and the number violation conditions due to the possibility of positive transfer from the L1 in the case of number agreement processing and the absence of this L1–L2 transfer possibility in the case of gender agreement processing. A final point important to bear in mind is that our participants are older than those of most previous studies, where undergraduate students have frequently been the population tested. Age-related changes in the distribution and indeed the timing of ERP components have been well documented (Kutas & Iragui, 1998). Therefore, to control for this factor, we also include in our study a control group of age-matched Spanish native speakers.

METHODS

Participants

Late Bilingual Group

Twenty-three highly proficient late learners of Spanish (14 women) participated in the experiment in exchange for a small sum. One participant was excluded from the statistical analyses due to too many eye blinks and muscular artifacts in the EEG. All participants were L1 speakers of English, with no history of neurological or psychiatric impairment and with normal or corrected-to-normal vision. All of them started to learn Spanish after the age of 20 and have had at least 12 years of immersion in a Spanish environment (mean = 22.1 years, range = 12–33 years). Their ages ranged from 35 to 60 years (mean = 48.6 years). All participants were right-handed, as assessed by an abridged version of the Edinburgh Handedness Inventory (Oldfield, 1971). The participants were interviewed by telephone and then in person by a trained native speaker of English and a trained native speaker of Spanish to evaluate their comprehension and fluency in both languages. All were judged to be native speakers of English and highly proficient in speaking and understanding Spanish. All participants filled in a self-rating questionnaire about their language background, their daily use of Spanish and English, and their proficiency in both languages. The results of this indicated that the participants use Spanish on a daily basis, are equally comfortable in most situations in both English and Spanish, and rate their competence in Spanish as being good to perfect (on a scale of 1 to 5, the average for speaking was 4, listening 4, reading 5, and writing 3). After the experiment, they also filled in a short grammar test, with a section on adjective agreement that included the experimental items. The results from this confirmed that off-line, the participants scored over 98% in the grammar test and were able to choose the correct form of the adjective in terms of gender and number agreement, at 98% accuracy.

Spanish Native-speaker Group

Twenty-three participants (10 women) participated in the experiment in exchange for a small sum. All of them were native speakers of Spanish, with no history of neurological or psychiatric impairment, and with normal or corrected-to-normal vision. Their ages ranged from 34 to 60 (mean = 48 years). All participants were right-handed, as assessed by an abridged version of the Edinburgh Handedness Inventory (Oldfield, 1971). After the experiment, they also filled in the short grammar test, with the section on adjective agreement that included the experimental items. The results from this confirmed that off-line, the participants scored 99% in the grammar test and were able to choose the correct form of the adjective in terms of gender and number agreement, at 99% accuracy. There was no statistically significant difference between the off-line results of the two groups. The participants in both groups were matched in terms of educational background and socioeconomic status.

Materials

The materials were the same as those used by Barber and Carreiras (2005). The details of these materials are as follows: Two lists of 120 experimental sentences each were generated to manipulate agreement at the beginning and in the middle of the sentence. The first list manipulated the article–noun agreement at the beginning of the sentence, whereas the second list manipulated the agreement relationship between the noun and its postmodifying adjective that appeared after the verb in the middle of the sentence (see examples below). The agreement condition is effectively the same in both sentence positions, but as this is the baseline for the ERP effects and is likely to differ in terms of amplitude from the first to the second sentence position, we consider these as six conditions for analysis purposes (a–f).

Manipulation at the beginning of the sentence:

(a) *Agreement*: El suelo está plano y bien acabado. (The$_{\text{masc.-sing.}}$ floor$_{\text{masc.-sing}}$ is flat and well finished.)
(b) *Gender Disagreement*: La suelo está plano y bien acabado. (The$_{\text{fem-sing}}$ floor$_{\text{masc.-sing}}$ is flat and well finished.)
(c) *Number Disagreement*: Los suelo está plano y bien acabado. (The$_{\text{masc.-plural}}$ floor$_{\text{masc.-sing.}}$ is flat and well finished.)

Manipulation at the middle of the sentence:

(d) *Agreement*: El suelo está plano y bien acabado. (The $_{\text{masc.-sing.}}$ floor $_{\text{masc.-sing}}$ is flat and well finished.)
(e) *Gender Disagreement*: El suelo está plana y bien acabado. (The $_{\text{masc.-sing}}$ floor $_{\text{masc.-sing}}$ is flat $_{\text{fem.-sing}}$ and well finished.)
(f) *Number Disagreement*: El suelo está planos y bien acabado. (The $_{\text{masc.-sing}}$ floor $_{\text{masc.-sing}}$ is flat $_{\text{masc.-plural}}$ and well finished.)

Between two- and four-filler words were included after the target adjective to avoid wrap-up effects on the target words. Assignment of sentences to conditions in each list was counterbalanced across participants. Thus, each sentence occurred three times across subjects, once in each condition, so that each subject only saw one form of each sentence. In addition, a list of 80 well-formed filler sentences was included. Fillers included nouns with opaque or irregular gender and adjectives with neuter gender to avoid strategies based on purely orthographic features. All the target items were medium to high-frequency words and contained between three and seven letters. The mean length and frequency of the words in each experimental condition were equivalent in each counterbalanced group. Word frequencies were obtained from the LEXESP database (Sebastián-Gallés, Martí, Carreiras, & Cuetos, 2000).

Procedure

Participants were seated comfortably in a darkened sound-attenuated chamber. All stimuli were presented on a high-resolution monitor that was positioned at eye level 80–90 cm in front of the participant. The words were displayed in light-gray lowercase Arial 36 against a black background. Participants performed a grammatical decision task—they were instructed to press one of two keys (L and S) on a keyboard to indicate whether the sentence was grammatically correct or not. For half of the participants, the right key (L) was used to signal the "yes" response and the left key (S) was assigned the "no" response. For the remaining participants, the order was reversed. The sequence of events in each trial is described as follows: First, a fixation point (*) appeared in the center of the screen and remained there for 700 msec. This fixation point was followed by a blank screen interval of 300 msec, then the sentence was displayed word by word. Each word appeared for 300 msec and was followed by a 300-msec blank interval. At the end of the sentence, a question mark cue to respond was presented and remained there up to a maximum of 2000 msec or until the participant's response. The intertrial interval varied randomly between 1000 and 1500 msec. Participants were asked to avoid eye movements and blinks if possible during the interval starting from the fixation point until response was given and were instructed to favor accuracy over speed in their responses. A practice session of five trials was given. The sentences were presented in three blocks of approximately 14 minutes each, with two short rest breaks. The whole experiment lasted approximately 90 minutes, including set-up time.

EEG Recording and Analyses

Scalp voltages were collected from 58 Ag/AgCl electrodes that were mounted in an elastic cap (10-10 system; ElectroCap International, Eaton, OH). Figure 1 shows the schematic distribution of the recording sites. Linked earlobes were used as reference. Eye movements and blinks were monitored with six further electrodes providing bipolar recordings of the horizontal and vertical EOG. Interelectrode impedances were kept below 10 KΩ. EEG was filtered with an analogue band-pass filter of 0.01–100 Hz, and a digital 20-Hz low-pass filter was applied before analysis. The signals were sampled continuously throughout the experiment with a sampling rate of 250 Hz.

Epochs of the EEG corresponding to 1500 msec after word onset presentation were averaged and analyzed. Baseline correction was performed using the average EEG activity in the 200-msec preceding the onset of the target stimuli as a reference signal value. Following baseline correction, epochs with artifacts were rejected, and trials that were not responded to correctly were not included in the analysis. This resulted in the exclusion of approximately 17% of the trials. Separate ERPs were formed for each of the experimental conditions, each of the subjects, and each of the electrode sites. Nine ROIs were computed out of the 58 electrodes, each containing the mean of a group of electrodes. These regions were as

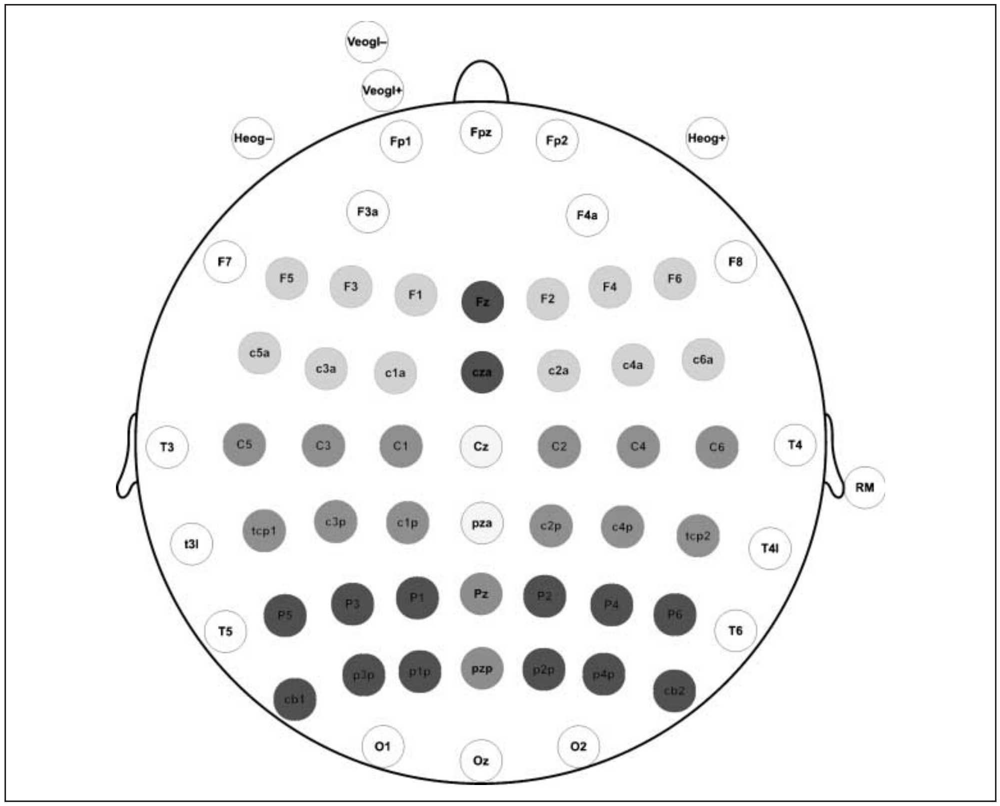

Figure 1. Schematic distribution of recording sites, indicating lateral and midline electrode groups.

follows (see electrode numbers in Figure 1): left anterior (F1, F3, F5, C1A, C3A, C5A), left medial (C1, C3, C5, C1P, C3P, TCP1), left posterior (P1, P3, P5, P1P, P3P, CB1), right anterior (F2, F4, F6, C2A, C2A, C2A), right medial (C2, C4, C6, C2P, C4P, TCP2), right posterior (P2, P4, P6, P2P, P4P, CB2), midline anterior (Fz, CZA), midline-medial (CZ,PZA), and midline posterior (PZ, PZP).

Mean amplitudes were obtained for different time windows. For each window, repeated measure ANOVAs were performed for the lateral electrode groups and for the midline electrode groups. The former included grammatical agreement (agreement, gender violation, and number violation), hemisphere (left/right), electrode region (anterior, medial, and posterior), and sentence position (beginning/middle) as the within-subject factors and group (native speakers/late learners) as the between-subject factor. The ANOVA for the midline groups included grammatical agreement (agreement, gender violation, and number violation), electrode region (anterior, medial, and posterior), and sentence position (beginning/middle) as the within-subject factors and group (native speakers/late learners) as the between-subject factor. Where appropriate, critical values were adjusted using the Greenhouse–Geisser (1959) correction for violation of the assumption of sphericity. In addition, post hoc Sidak contrasts were performed after interactions or main effects of grammatical agreement to control for type 1 error in multiple comparisons. Effects for the hemisphere factor, for the electrode region factor, and for the sentence position factor are only reported when they interact with the experimental manipulations.

RESULTS

Behavioral Results

RTs were not analyzed, as the response in the grammatical judgment task was delayed and participants were instructed to favor accuracy over speed. In terms of error rates, the mean percentage of correct answers in each condition for each group can be seen in Table 1. The overall mean percentage of correct answers in the native-speaker group was 95.7%. The overall mean percentage of correct answers for the late learner group was 88.7%. ANOVA on the error rates including grammatical agreement (agreement, gender violation, and number violation) and sentence position (beginning/middle) as within-subject factors and group (native speakers/late learners) as between-subject factor revealed a reliable interaction of grammatical agreement, sentence position, and group, $F(2, 86) = 9.76$, $p < .001$, $\varepsilon = .775$.

For the group of native speakers, pairwise comparisons revealed significant differences in the second sentence position: They made more errors in the number violation condition than in both the agreement, $F(1, 22) = 6.83$, $p < .05$, and the gender violation condition, $F(1, 22) =$

Table 1. Mean Percentage Values of Correct Responses to the Grammatical Task for Each Group in the Experimental Conditions

	Native Speakers		*Late Learners*	
	Position 1	*Position 2*	*Position 1*	*Position 2*
Agreement	95.76 (0.88)	97.38 (1.19)	94.40 (.90)	90.36 (1.22)
Gender	96.84 (2.83)	97.50 (1.40)	75.62 (2.90)	86.04 (1.43)
Number	93.04 (1.64)	93.47 (2.10)	92.32 (2.15)	93.77 (1.66)

Values are presented as mean (*SE*).

8.31, $p < .05$. For the group of late learners, pairwise comparisons revealed that in the first sentence position, they made more errors in the gender violation condition than in both the number violation condition, $F(1, 21) = 30.45$, $p < .001$, and the agreement condition, $F(1, 21) = 37.30$, $p < .001$. This was the same for the second position: They made more errors in the gender violation condition than in both the number violation condition, $F(1, 21) = 23.89$, $p < .001$, and the agreement condition, $F(1, 21) = 8.07$, $p < .05$.

ERP Results

ERP grand averages time locked to the onset of the target words are represented in Figures 2, 3, 4, and 5 over representative recording sites, showing the three conditions: agreement, gender violation, and number violation. Figure 2 refers to the native-speaker group—first sentence position, Figure 3 to the native-speaker group—second sentence position, Figure 4 to the late learners group—first sentence position, and Figure 5 to the late learner group—second sentence position.

Native Speakers

Visual inspection of both figures reveals clear differences in the responses to the gender and number disagreement conditions with respect to the agreement conditions. Between 300- and 450-msec disagreement waves for gender and number violations display larger negative amplitudes than the agreement condition, particularly on the left side of the scalp, fitting with the aforementioned LAN effect. There are no apparent differences between the gender and the number conditions or between the first and the second sentence positions. Between 500 and

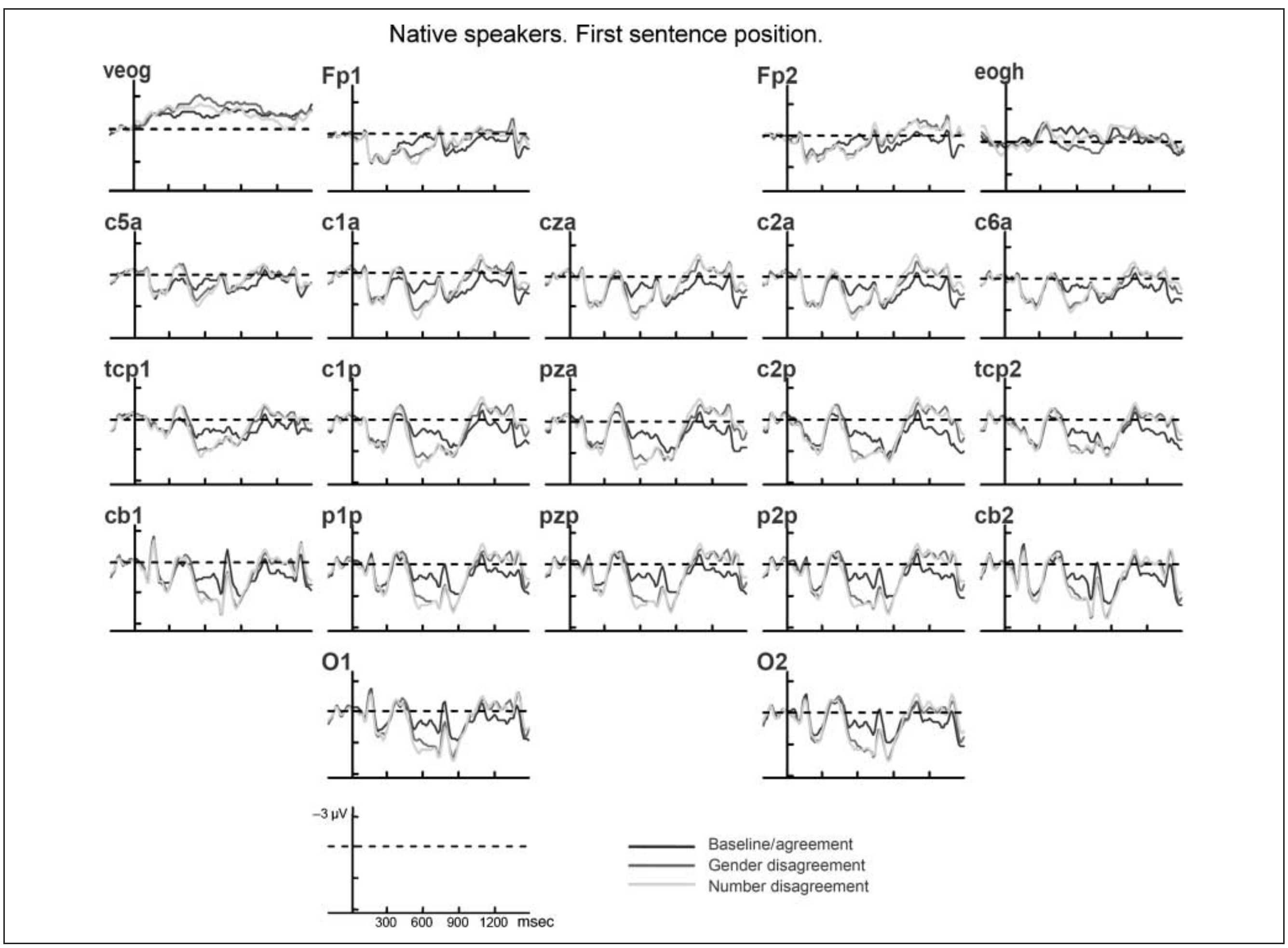

Figure 2. Grand average ERPs corresponding to the grammatical agreement manipulation (first sentence position) in native speakers.

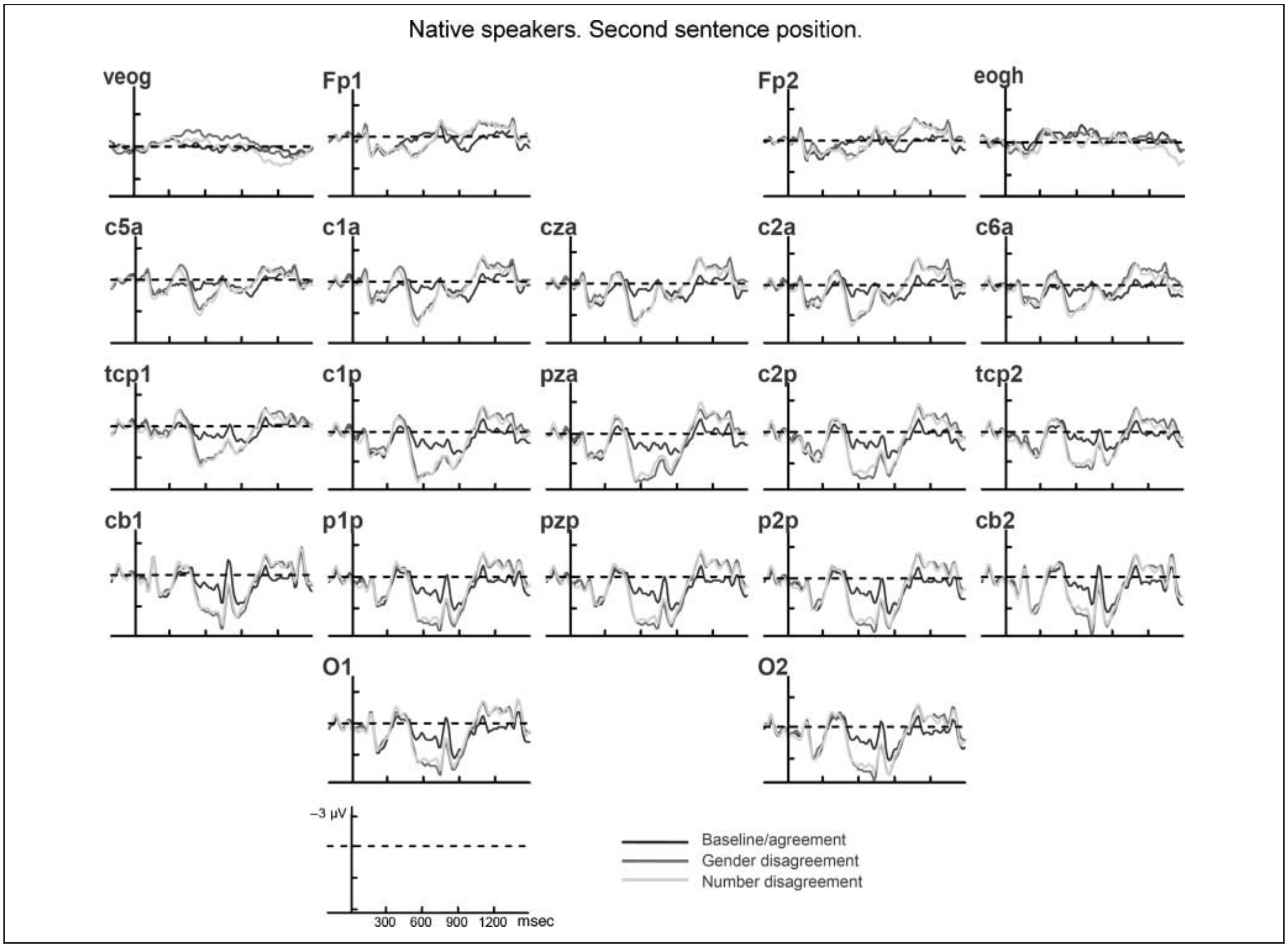

Figure 3. Grand average ERPs corresponding to the grammatical agreement manipulation (second sentence position) in native speakers.

900 msec, in both sentence positions, there is then a typically widely distributed P600 effect consisting in larger positive amplitudes for both disagreement conditions compared with the agreement condition. Again, there are no apparent differences between the gender and the number conditions or between sentence positions. Finally, in both figures, a sustained negativity can be observed between around 1000 and 1300 msec, with more negative amplitudes for the disagreement conditions, again with no apparent differences between the violation conditions and the sentence position conditions.

Late Learner Group

Visual inspection of these figures again reveals clear differences in the responses to the gender and the number disagreement conditions with respect to the agreement conditions. However, in contrast to the native-speaker group, there are differences between the violation conditions and between the first and the second sentence positions. First, between 300 and 450 msec, in the first sentence position, disagreement waves for both gender and number violations display larger negative amplitudes than the agreement condition, particularly on the left side of the scalp, although this negativity effect appears to be more widely distributed than in the native-speaker group, particularly in the case of number disagreement. In the second sentence position, this effect is absent. Between 500 and 900 msec, in both sentence positions, there is a widely distributed P600 effect consisting in larger positive amplitudes for both disagreement conditions compared with the agreement condition, but in contrast to the native-speaker group, the number violation condition appears to be more positive than the gender violation condition. Finally, in both figures, a sustained negativity can again be observed between around 1000 and 1300 msec, with more negative amplitudes for the disagreement conditions. Again, this effect appears to be greater for the number than for the gender agreement violation.

Following visual inspection of the grand averaged waveforms, omnibus ANOVAs were carried out separately for lateral and midline electrode groups in three different temporal windows on the basis of calculations of mean amplitudes: between 300 and 450 msec for the LAN effect, between 500 and 900 msec for the P600 effect, and between 1000 and 1300 msec for the late negativity effect.

300–450 msec: early negativity.

LATERAL ELECTRODE GROUPS. ANOVA revealed a near to significance four-way interaction including agreement, hemisphere, electrode, and group, $F(4, 172) = 2.25$, $p = .08$, $\varepsilon = 0.765$.

MIDLINE ELECTRODE GROUPS. There were no reliable effects in this time window in the midline electrode groups.

Although the four-way interaction (lateral electrode groups) was only near significance, visual inspection clearly shows that effects for native and L2 learners are similar in the first sentence position whereas they are quite different in the second sentence position. Therefore, to better characterize the data, we opted to carry out separate analyses for natives and L2 learners.

Native-speaker group.

LATERAL ELECTRODE GROUPS. ANOVA in this window yielded an interaction of grammatical agreement, hemisphere, and electrode region, $F(4, 88) = 2.91$, $p < .05$, $\varepsilon = 0.727$. Both the gender and the number violation conditions were more negative than the agreement condition but there were no significant differences between the violation conditions. Pairwise comparisons are shown in Table 2.

MIDLINE ELECTRODE GROUPS. No significant differences were obtained.

Late learner group.

LATERAL ELECTRODE GROUPS. ANOVA in this window yielded a four-way interaction of the factors grammatical agreement, hemisphere, electrode region, and sentence position, $F(4, 84) = 2.95$, $p < .05$, $\varepsilon = 0.811$. Significant effects of grammatical agreement were only in the first sentence position, with the gender and the number violation conditions more negative than the agreement condition and no significant differences between the violation conditions. Pairwise comparisons are shown in Table 3.

MIDLINE ELECTRODE GROUPS. No significant differences were obtained.

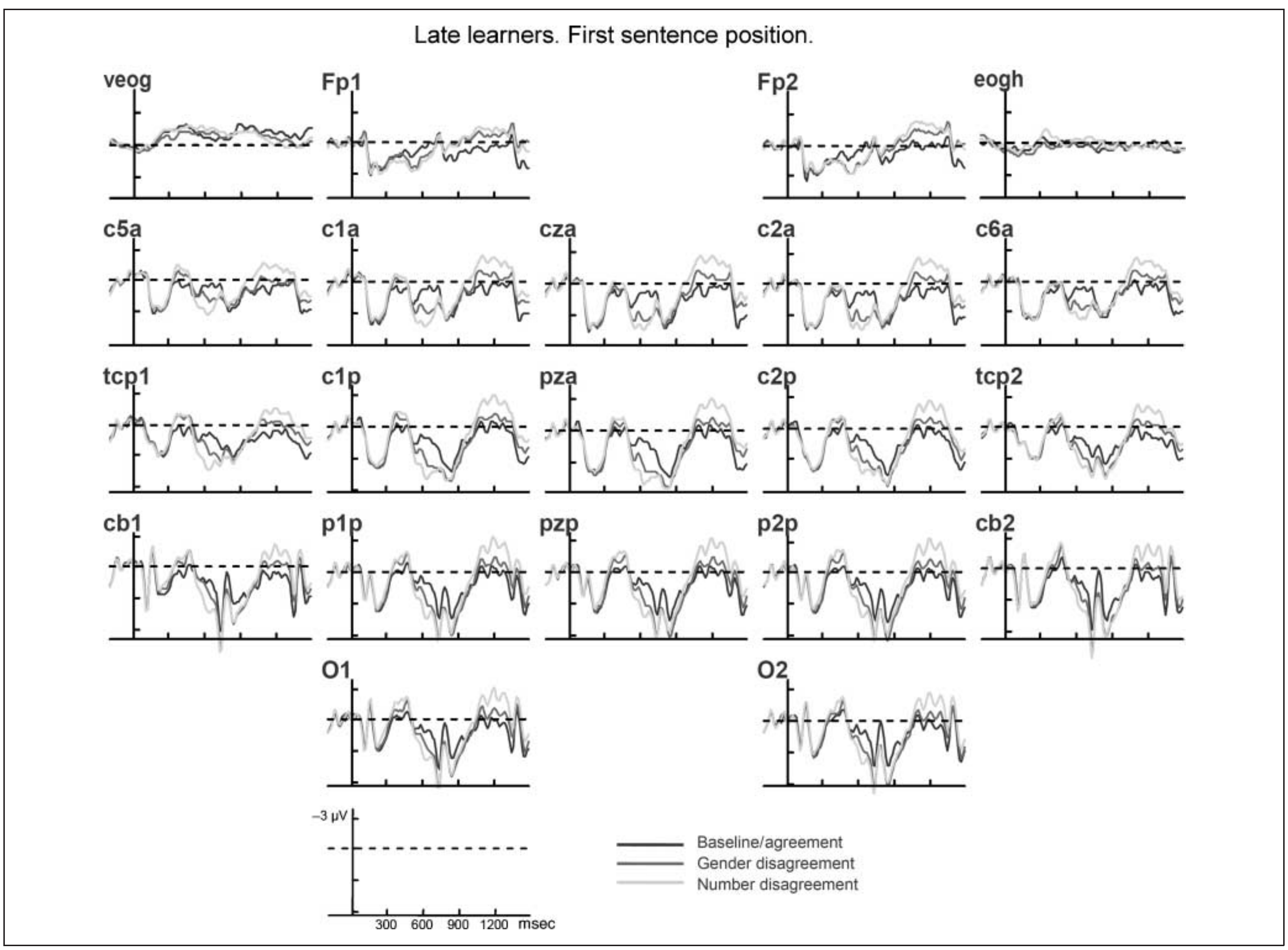

Figure 4. Grand average ERPs corresponding to the grammatical agreement manipulation (first sentence position) in late learners.

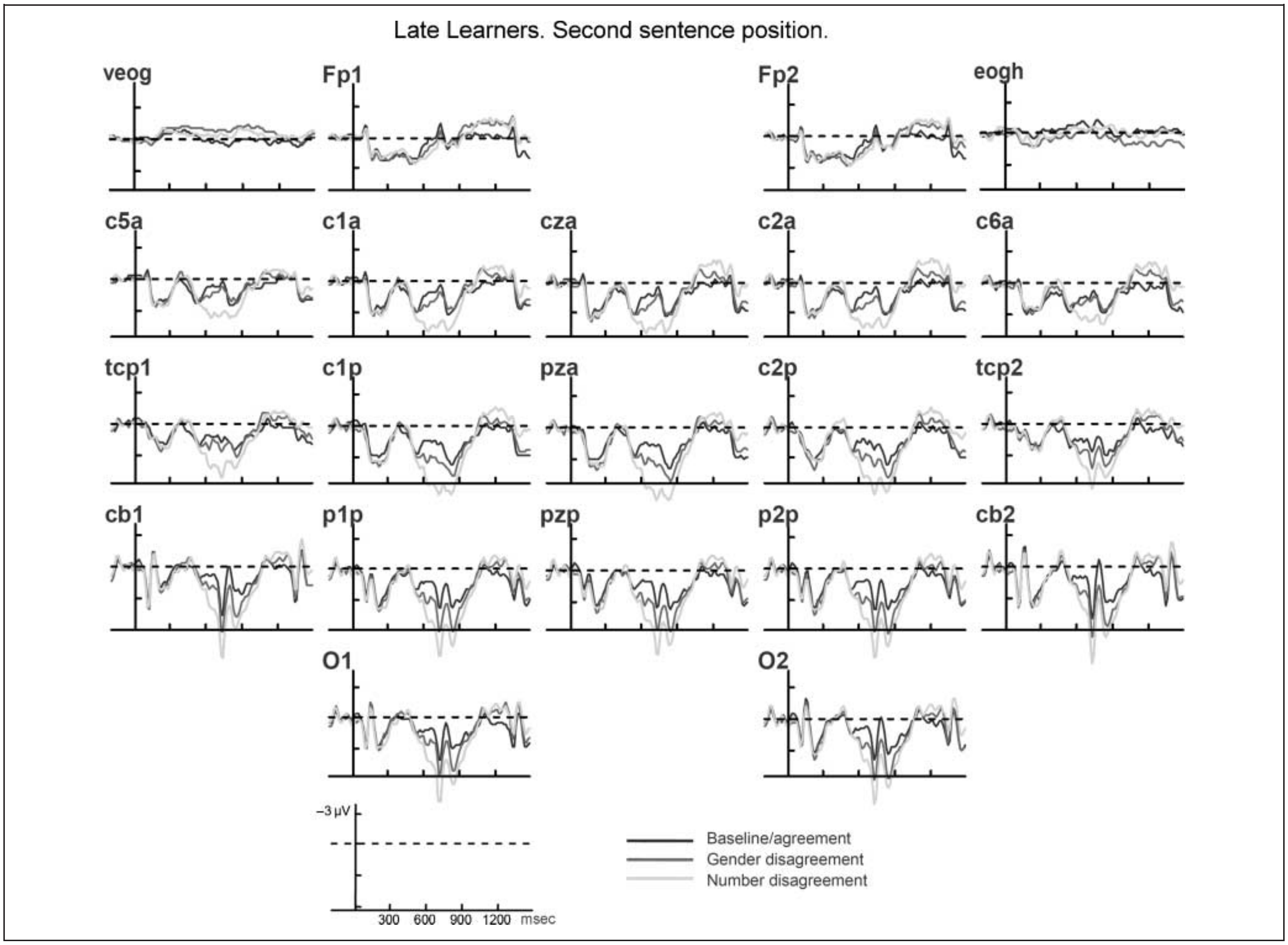

Figure 5. Grand average ERPs corresponding to the grammatical agreement manipulation (second sentence position) in late learners.

500–900 msec: P600.

Lateral electrode groups. ANOVA yielded a reliable interaction of grammatical agreement and group, $F(2, 86) = 4.15$, $p < .05$, $\varepsilon = .989$. Paired contrasts (see Table 2) revealed that, in the case of native speakers, both violation conditions were more positive than the agreement condition but there were no differences between the gender and the number violations. In contrast, in the case of the late learners (see Table 3), although both violation conditions were more positive than the agreement condition, there were also significant differences between gender and number violation, with number showing larger positivity than gender.

Midline electrode groups. ANOVA yielded a reliable interaction of grammatical agreement and group, $F(2, 86) = 4.63$, $p < .05$, $\varepsilon = 0.971$. Paired contrasts revealed that in the case of the native-speaker group (see Table 2), the violation conditions were more positive than the agreement condition, but there were no differences between gender and number violation. In contrast, in the case of the late learners (see Table 3), although both violation conditions were more positive than the agreement condition, there were also significant differences between gender and number violation with number showing larger positivity than gender.

1000–1300 msec: late negativity.

Lateral electrode groups. ANOVA yielded a reliable interaction of grammatical agreement, hemisphere, and electrode region, $F(4, 172) = 4.43$, $p < .005$, $\varepsilon = .828$, but yielded no significant effects of group.

Midline electrode groups. ANOVA yielded a main effect of grammatical agreement, $F(2, 86) = 29.35$, $p < .001$, $\varepsilon = 0.867$, but no interaction with the factor group. Although there are no reliable effects of group in this time window, visual inspection of the figures shows a clearly different pattern of results for the natives and the late learners in terms of effects in the first and the second sentence positions. Therefore, we again opted to carry out separate analyses for the two groups.

Native-speaker group.

Lateral electrode groups. ANOVA yielded a reliable interaction of grammatical agreement, hemisphere, and electrode

Table 2. Statistics for the Main Contrasts (Agreement vs. Gender Violation, Agreement vs. Number Violation) of the Significant Interactions Obtained for the Group of Native Speakers Across the Windows of Analysis (*df* = 1, 22)

	Gender						Number					
Native Speakers		*F*	*p*		*F*	*p*		*F*	*p*		*F*	*p*
300–450 msec												
Lateral analysis												
Agreement × Hemisphere × Electrode Region	LA	24.06	<.001	RA	5.69	.07	LA	8.97	<0.5	RA	3.69	.18
	LM	17.48	<.01	RM	6.58	.05	LM	9.89	<0.5	RM	4.12	.15
	LP	2.38	.35	RP	<1		LP	3.40	.21	RP	<1	
500–900 msec												
Lateral analysis												
Agreement		38.01	<.001					31.82	<.001			
Midline analysis												
Agreement		47.80	<.001					32.56	<.001			
1000–1300 msec												
Lateral analysis												
Agreement × Hemisphere × Electrode Region	LA	14.21	<.01	RA	34.59	<.01	LA	18.82	<.01	RA	37.23	<.01
	LM	13.74	<.01	RM	27.45	<.01	LM	12.20	<.01	RM	19.30	<.01
	LP	12.04	<.01	RP	16.83	<.01	LP	14.06	<.01	RP	16.41	<.01
Midline analysis												
Agreement		27.41	<.001					22.61	<.001			

region, $F(4, 88) = 3.55, p < .05, \varepsilon = .738$. Paired contrasts (see Table 2) revealed that the gender and the number violation conditions were more negative than the agreement condition, but there were no differences between the violation conditions.

Midline electrode groups. ANOVA yielded a reliable main effect of grammatical agreement, $F(2, 44) = 17.54, p < .001, \varepsilon = 0.910$. Paired contrasts (Table 2) indicated that the gender violation condition was more negative than the agreement condition and the number violation condition was also more negative than the agreement condition.

Late learner group.

Lateral electrode groups. ANOVA yielded a marginal interaction of grammatical agreement, hemisphere, and sentence position, $F(2, 42) = 3.08, p < .06, \varepsilon = .961$. Paired contrasts (see Table 3) revealed that the number violation condition was more negative than the gender violation condition and the agreement condition.

Midline electrode groups. ANOVA yielded a reliable interaction of grammatical agreement and sentence position, $F(2, 42) = 3.45, p < .05, \varepsilon = 0.927$. Paired contrasts (see Table 3) indicated that in the first sentence position, the number violation condition was more negative than the agreement condition and the number violation condition was also more negative than the gender violation condition. There were no significant effects in the second sentence position.

Summary of results

For both groups, the violations of grammatical agreement in the sentences presented resulted in three effects, two of which have been previously reported as related to syntactic processing. There were differences between the two groups, however, in the amplitude and distribution of these effects in the different conditions. For the native-speaker group, the pattern was the same in both sentence positions and consisted in, first, a left-biased negativity around 300–500 msec that corresponds to previous reports of LAN effects, second, a generally distributed positive deflection between 500 and 900 msec corresponding to the P600, and third, a late negativity between 1000 and 1300 msec that has been previously reported as possibly relating to working memory/task demands (Sabourin & Stowe, 2008). There were no

Table 3. Statistics for the Main Contrasts (Agreement vs. Gender Violation, Agreement vs. Number Violation, and Gender Violation vs. Number Violation) of the Significant Interactions Obtained for the Group of Late Learners Across the Windows of Analysis (*df* = 1, 21)

	Gender						*Number*						*Gender <> Number*		
Late Learners		*F*	*p*		*F*	*p*		*F*	*p*		*F*	*p*		*F*	*p*
300–450 msec															
Lateral analysis															
Agreement × Hemisphere × Electrode Region × Position															
Position 1	LA	2.72	.30	RA	<1		LA	3.06	.25	RA	<1				
	LM	6.92	<.05	RM	1.77	.48	LM	12.20	<.01	RM	4.05	.16			
	LP	3.87	.17	RP	1.58	.52	LP	13.52	.005	RP	10.75	.01			
Position 2	LA	<1		RA	<1		LA	<1		RA	1.96	.44			
	LM	<1		RM	<1		LM	1.10	.66	RM	2.10	.41			
	LP	<1		RP	<1		LP	<1		RP	<1				
500–900 msec															
Lateral analysis															
Agreement		14.88	<.01					53.94	<.001					16.70	<.01
Midline analysis															
Agreement		21.82	<.001					59.31	<.001					16.36	<.01
1000–1300 msec															
Lateral analysis															
Agreement × Hemisphere × Position															
Position 1	LH	10.28	<.05				LH	29.2	<.001				LH	16.53	<.01
	RH	5.38	.40				RH	29.13	<.001				RH	19.25	<.001
Position 2	LH	2.70	.30				LH	4.46	.13				LH	<1	
	RH	3.54	.20				RH	7.89	<.05				RH	2.87	.28
Midline analysis															
Agreement × Position															
Position 1		5.78	.07					30.09	<.001					20.18	<.01
Position 2		2.07	.41					5.67	.07					2.35	.36

significant differences between the gender and the number violation conditions in any of these effects.

For the late learner group, in contrast, there were important differences between the two sentence position manipulations and between the gender and the number violations. The early negativity was only present in the first sentence position violations, and although there are no statistically significant differences between the violation conditions in this time window, the effect is more widely distributed for the number violation condition, reaching significance in left posterior and right posterior scalp areas as well as in the left central area, which is the only area significant for the gender violation. We will analyze this time window in more detail below to try to identify these effect differences with more precision. Between 500 and 900 msec in both sentence positions, there was then a broadly distributed positive deflection that corresponds well to descriptions of the P600. Here again, however, there were differences between the violation conditions, with number violation resulting in significantly greater positivity than gender violation, in contrast to the results from the native-speaker group. In addition, the late learner groups also showed a late negativity, again with differences between the violation conditions, with the number violation condition more negative than the other conditions.

Early negativity—a LAN-type effect? Although the distribution of the early negativity in the control group of native speakers fits fairly closely with most reported LAN effects, the distribution of this effect in the late learners (in the first sentence position only) is different from the native speakers—particularly in the case of number violation. Visual inspection of the grand averages of both groups also reveals differences between the two groups in the onset and duration of the early negativity. Although the analysis window (300–450 msec) was consistent for purposes of comparison between the groups, it is clear that the effects develop across a slightly longer window in the case of the late learners. These differences can be more clearly appreciated in Figure 6 in which there is a comparison between groups of the difference waves for each violation condition compared with the agreement condition as well as topographical maps showing the development of the effects in each group.

To examine these differences more closely, it was decided to carry out amplitude analyses on 20-msec time windows from 300 msec onward to locate and to compare the onset of this early negativity evoked by the gender and number violations. Repeated measures ANOVAs were again performed, one for the lateral electrode groups and one for the midline electrode groups. The results are as follows:

340–360 MSEC. As can be seen in Table 4, this is the first window in which there are statistically significant effects for either group. For the native-speaker group, the ANOVA in this window indicated a three-way interaction between grammatical agreement, hemisphere, and electrode region, $F(4, 88) = 5.43, p < .005, \varepsilon = 0.606$. For the late learner group, the analyses yielded a four-way interaction between grammatical agreement, hemisphere, sentence position, and electrode region, $F(4, 84) = 2.91$, $p < .05$, $\varepsilon = 0.853$. Significant contrasts for this and the other short time windows can be seen in Table 4. In this time window, at the first sentence position, both late learners and native speakers have a significant LAN-type effect for number, but only the native speakers have a LAN effect for gender. At the second sentence position, the contrast is sharper, as the late learners show no significant effects but the native speakers have LAN-type effects for both gender and number.

400–420 MSEC. For the late learner group, this is the first window in which the LAN-type effect for gender violation reaches significance. For this group, ANOVA in this window yielded a significant four-way interaction between grammatical agreement, hemisphere, sentence position, and electrode region, $F(4, 84) = 2.71, p = .05, \varepsilon = 0.745$.

Thus, it can be seen that for the native-speaker group, both the gender and the number violation conditions elicit a LAN that becomes statistically significant at around 340–360 msec with no effects of sentence position. However, for the late learner group, in the same time window, this effect is restricted to the first sentence position and to the number violation condition. For this group in a later time window, however, from 400 to 420 msec, there is a statistically significant left-biased negativity effect of the gender violation condition as well as the effect of number violation that has extended beyond the left medial area to posterior and right hemisphere areas.

Summary of results

In general terms, the ERP pattern of response of both groups to the morphosyntactic violations is fairly similar. The biphasic ERP response of early negativity followed by P600, which has been frequently observed in previous studies, is repeated here. However, there are interesting differences between the two groups in terms of the onset and amplitude of the response to gender and number violations. First, the early negativity is only detectable for the late learners in the first sentence position, which is not the case for the native speakers. Second, although this left-biased negativity response to number agreement violation is at the onset quite similar in the two groups, in the case of the late learner group it has a longer duration than for the native speakers and appears to "spread" to posterior and right hemisphere scalp areas. Also, the response to gender agreement violation is delayed in the case of the late learners with respect to the native speakers as this left-biased negativity has an onset that is around 60 msec later in the case of the late learners. In the later time windows, although both groups have clear P600 effects, again the native speakers show no effects of sentence position and no differences between the gender and the number violation disagreement conditions, whereas for the late learner group, the number violation condition is significantly more negative than the gender violation condition. Finally, although the late negativity is observed in both groups, again there are differences for the late learner group with the number disagreement condition eliciting stronger effects than the gender violation condition.

DISCUSSION

According to the CPH, there are maturational constraints on the learning of L2 or other languages. This premise is incorporated in the failed functional features hypothesis (Hawkins & Chan, 1997), which posits that, for late learners of an L2, only language features that are already instantiated in the L1 will be available and transferred from the L1 to the L2. However, recent evidence indicates that although age is certainly an important factor in L2 acquisition, it is not the only factor of relevance. In studies of the representation and processing of late-learned L2, it is also important to take into consideration other factors such as the level of proficiency attained. In contrast to the failed functional features hypothesis, the full transfer/full access view (Schwartz & Sprouse, 1996)

Figure 6. Early negativity—difference waves and maps for each violation condition compared with the agreement condition.

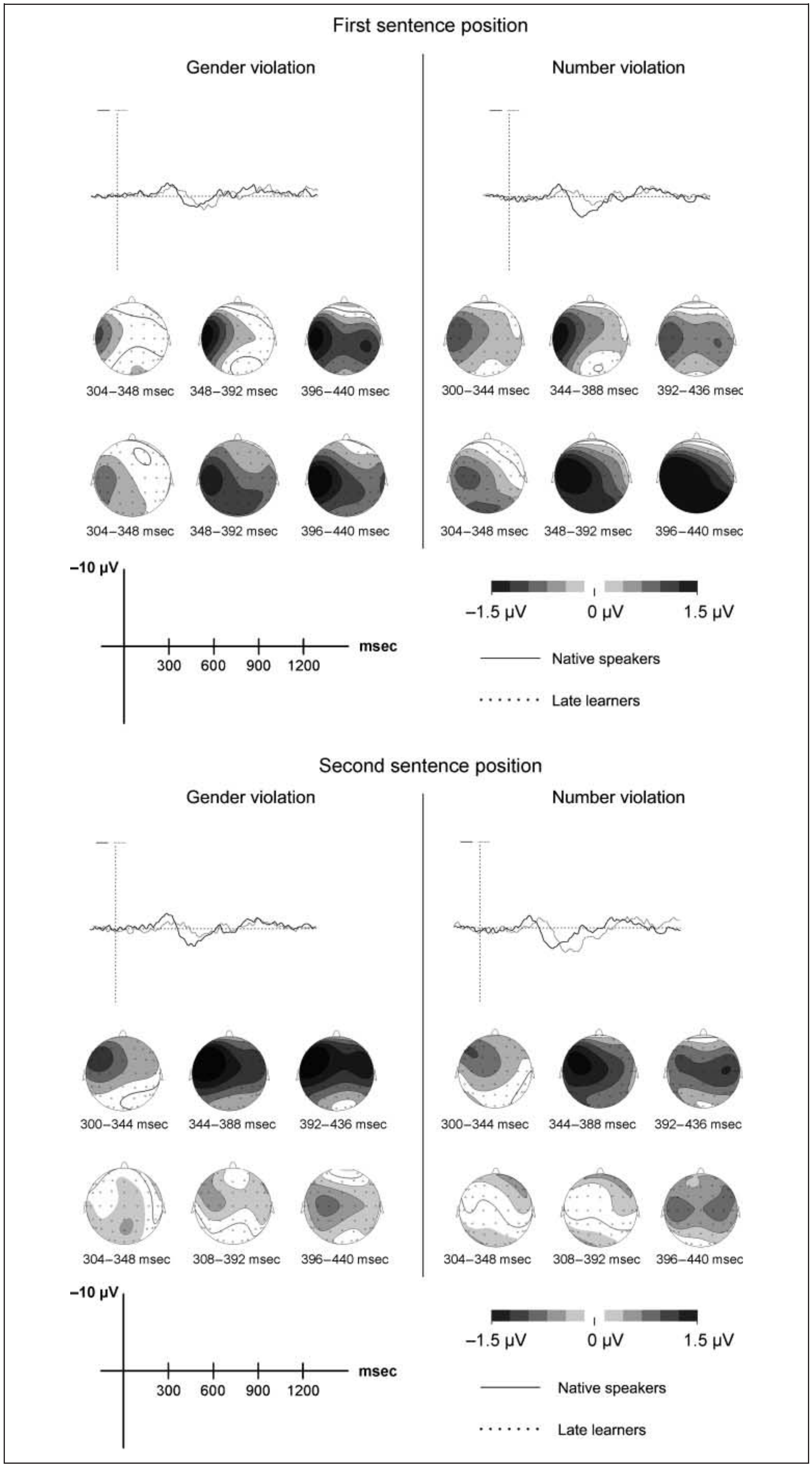

Table 4. Significant Contrasts for Lateral Electrodes across 20-msec Time Windows from 300 to 440 msec

	Native Speakers				*Late Learners*			
	P1		*P2*		*P1*		*P2*	
	G	*N*	*G*	*N*	*G*	*N*	*G*	*N*
300–320	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
320–340	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
340–360	LA**	LA**	LA**	LA**	*ns*	LM*	*ns*	*ns*
	LM**	LM**	LM**	LM**				
360–380	LA**	LA**	LA**	LA**	*ns*	LM*	*ns*	*ns*
	LM**	LM**	LM**	LM**		LP*		
380–400	LA**	LA**	LA**	LA**	*ns*	LM**	*ns*	*ns*
	LM**	LM**	LM**	LM**		LP*		
400–420	LM*	*ns*	LA**	*ns*	LA*	LM** RM*	*ns*	*ns*
			LM*		LM*	LP** RP**		
420–440	LA*	*ns*	LA*	*ns*	LM*	LM** RM*	*ns*	*ns*
	LM*		LM*			LP* RP*		

$^{*}p < .05$.
$^{**}p < .01$.

argues that although the starting point or initial state of the L2 is indeed the L1 grammar, with continued exposure to the L2 and high degrees of proficiency not only will these language features or processing routines be available for processing in the L2, but those of the L2 not present in the L1 can be acquired. Our study tested these different assumptions by recording the electrical brain activity of a group of English L1 participants who began to learn Spanish after the age of 20 years, who have lived in the L2 environment for an average of 22 years, and who have attained high levels of proficiency in Spanish. When processing Spanish sentences containing violations of grammatical gender agreement (a feature not present in their L1) and number agreement (a feature that is computed in the L1), the ERP patterns of these late learners are generally similar to those of an age-matched native-speaker group, although not identical. In the first sentence position, the late learners show the classic pattern of left-biased negativity followed by P600 for both gender and number violation. Thus, at least for this determiner–noun construction, both gender and number agreement would seem to be processed in a native-like manner, as argued by the full transfer/full access position. There are, however, some differences between the effects for the gender and those for the number violation manipulations, which is not the case for the native-speaker group. These first position differences consist in an earlier onset of the negativity effect for number as opposed to gender violation as well as a wider distribution of the negativity response to number violation across the time window and greater amplitude of the P600 effects for the number violation condition compared with the gender violation condition. However, quantitative differences between gender and number processing are in fact predictable by the full transfer/full access model, according to which these differences could be due to the fact that, as number is a grammatical feature in the L1 (English), it is likely to be acquired earlier and processed in more "depth" than grammatical gender, which does not exist in the L1. The differences in the onset of the early negativity and in the amplitude of the P600 for gender and number processing, therefore, possibly reflect some differences in the speed and depth of processing of these features even at high levels of proficiency, with number violations processed faster and in more depth than gender violations. In the same way, late learners are showing more errors in judging the acceptability of sentences where gender is violated (compared with number violation), indicating that they have more problems with this feature than with number violation.

In the second sentence position manipulation, the differences between the groups are somewhat greater, whereas the native speakers show a similar negativity + P600 pattern to the first position manipulation, with no differences in amplitude between the experimental conditions. For the late learners, there are no significant effects in the early time window, although there is again a clear P600 effect with greater amplitudes for the number violation condition than for the grammatical gender condition and a late negativity effect with, again, greater amplitudes for number than for gender agreement manipulations. There are a number of possible explanations for this difference between the first and the second sentence positions. One explanation may lie in capacity limitation accounts of L2 acquisition and processing, which attribute the lack of native-like processing patterns often observed in these studies to the increased cognitive load involved in processing the L2, particularly in terms of working memory costs (McDonald, 2006). Processing of the first position grammatical agreement violations presumably involves a much lower cost, firstly because the determiner and the noun are adjacent and second because they are within the same phrase. A different case, however, is that of the second sentence position. Here the grammatical feature-checking involves a greater distance with an intervening element—the verb—and it is also an across-phrase violation. Even for native speakers, the higher cost of processing agreement with intervening elements has been well documented (Deutsch & Bentin, 2001). It is also possible that as determiner–noun agreement (the first position manipulation) is sometimes computed in the L1 English but adjective–noun agreement (the second position manipulation) is not, then the processing routines required for the first position violation are already available in some form to late learners whereas the processing mechanisms necessary for the second position violations are not. This would be more in line with the failed functional features hypothesis view. With the present design, it is not possible to confirm or to reject either of these possible explanations for the lack of a left negativity effect for the late learners in the second sentence position.

As well as the two syntax-related components, in both sentence positions, there was also a late negativity between 1000 and 1300 msec, which has been previously described as related to working memory and task demands (Sabourin et al., 2006; Ruchkin, Grafman, Cameron, & Berndt, 2003). Again, in this effect, there are differential responses for number and gender violations. In the first sentence position, the effect for number violation is greater than that for gender violation. In the second sentence position, only the number violation condition is significant. As to the significance of this effect, Sabourin and Stowe (2008) have suggested two possible explanations for this type of effect. In Sabourin and Stowe (2004), they suggested that this type of late negativity, which was more frontally distributed in their studies, could be related to the effort of maintaining the ungrammaticality in memory until a delayed grammaticality decision is made at the end of the sentence. More recently, in a study of gender agreement in late bilinguals (Sabourin & Stowe, 2008), they proposed that the late negativity could index a processing resource "used by an L2 group who are proficient enough to recognize that there is an error, but cannot easily make use of native processing routines to deal with the ungrammaticality." In the case of our experiment, only the first of these explanations could apply. The response and the final decision is delayed in both sentence positions, but although the late negativity effect is significant for number violations in both sentence positions, it is only significant for gender violations in the first sentence position, where there was also an early negativity effect for gender violation. Thus, this late negativity could indeed be task related, and index memory load where the ungrammaticality is detected and maintained in memory until the response is made.

Our results in general, therefore, would seem to argue that although AoA is a relevant factor in L2 processing, there are indeed other critical factors that have to be taken into account. The first of these is proficiency, as our results add to the growing body of evidence indicating that native-like language processing in late learners is possible at high levels of proficiency under some conditions. As to the question of language transfer and the availability of language features not originally present in the L1, according to our results the late learners are processing not only number—a grammatical feature already present in the L1—but grammatical gender, which is not a feature codified in their L1. In this sense, our results give more support to the full transfer/full access account than to the failed functional features hypothesis. However, the differential results for number and gender, with number violations clearly producing stronger effects, and for the two sentence positions could indicate the importance and lasting effects of L1 feature transfer even at advanced proficiency levels and the relevance of other cognitive aspects of L2 processing, such as working memory demands and resources.

Acknowledgments

This research was partially supported by grant AP2003-1787 from the Spanish Ministry of Science and Innovation to the first author, grants SEJ2006-09238/PSIC and CONSOLIDER-INGENIO2010 CSD2008-00048 to Manuel Carreiras, and grant SEJ2007-67364 from the Spanish Ministry of Science and Innovation to Horacio Barber. Marta Vergara was supported by a postdoctoral contract under the Spanish Fullbright-MEC commission.

Reprint requests should be sent to Margaret Gillon Dowens, Division of English Studies, Centre for Research in Applied Linguistics, University of Nottingham Ningbo, 199 Taikang East Rd., Ningbo 315100, China, or via e-mail: Margaret.Dowens@nottingham.edu.cn.

REFERENCES

Abutalebi, J., Cappa, S. F., & Perani, D. (2001). The bilingual brain as revealed by functional neuroimaging. *Bilingualism: Language and Cognition, 4,* 179–190.

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience, 17,* 137–153 (in Spanish).

Bialystok, E., & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 161–181). Mahwah, NJ: Lawrence Erlbaum Associates.

Birdsong, D. (1999). Introduction: Whys and why nots of the critical period hypothesis for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 1–22). Mahwah, NJ: Lawrence Erlbaum Associates.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning, 56,* 9–49.

Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics, 27,* 3–42.

Clahsen, H., & Felser, C. (2006b). How native-like is non-native language processing? *Trends in Cognitive Sciences, 10,* 564–570.

Coulson, S., King, J. W., & Kutas, M. (1998). ERPs and domain specificity: Beating a straw horse. *Language and Cognitive Processes, 13,* 653–672.

DeKeyser, R., & Larsen-Hall, J. (2005). What does the critical period really mean? In J. F. Kroll & A. M. D. B. Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*. Oxford: Oxford University Press.

Deutsch, A., & Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: Evidence from ERPs and eye movements. *Journal of Memory and Language, 45,* 200–224.

Eubank, L., & Gregg, K. R. (1999). Critical periods and (second) language acquisition: Divide et impera. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 65–99). Mahwah, NJ: Lawrence Erlbaum Associates.

Frenck-Mestre, C., Osterhout, L., McLaughlin, J., & Foucart, A. (2008). The effect of phonological realization of inflectional morphology on verbal agreement in French: Evidence from ERPs. *Acta Psychologica, 128,* 528–536.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1219–1248.

Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research, 31,* 45–63.

Friederici, A. D., Steinhauer, K., & Frisch, S. (1999). Lexical integration: Sequential effects of syntactic and semantic information. *Memory & Cognition, 27,* 438–453.

Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences, U.S.A., 99,* 529–534.

Greenhouse, S., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 32,* 95–112.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two native speakers in one person. *Brain and Language, 36,* 3–15.

Gross, M., Say, T., Kleingers, M., Clahsen, H., & Münte, T. F. (1998). Human brain potentials to violations in morphologically complex Italian words. *Neuroscience Letters, 241,* 83–86.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience, 15,* 883–899 (in Dutch).

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia, 38,* 1531–1549.

Hagoort, P., Wassenaar, M., & Brown, C. M. (2003). Syntax-related ERP effects in Dutch. *Cognitive Brain Research, 16,* 38–50.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research, 30,* 251–266.

Hasting, A. S., & Kotz, S. A. (2008). Speeding up syntax: On the relative timing and automaticity of local phrase structure and morphosyntactic processing as reflected in event-related brain potentials. *Journal of Cognitive Neuroscience, 20,* 1207–1219.

Hasting, A. S., Kotz, S. A., & Friederici, A. D. (2007). Setting the stage for automatic syntax processing: The mismatch negativity as an indicator of syntactic priming. *Journal of Cognitive Neuroscience, 19,* 386–400.

Hawkins, R., & Chan, C. Y. (1997). The partial availability of universal grammar in second language acquisition: The "Failed functional features hypothesis." *Second Language Research, 13,* 187–226.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints on SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 539–588). Malden, MA: Blackwell.

Jeong, H., Sugiura, M., Sassa, Y., Haji, T., Usui, N., Taira, M., et al. (2007). Effect of syntactic similarity on cortical activation during second language processing: A comparison of English and Japanese among native Korean trilinguals. *Human Brain Mapping, 28,* 194–204.

Johnson, J. S., & Newport, E. L. (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition, 39,* 215–258.

King, J. W., & Kutas, M. (1995). Do the waves begin to waver? ERP studies of language processing in the elderly. In P. A. Allen & T. R. Bashore (Eds.), *Advances in psychology* (pp. 314–344). Amsterdam: North-Holland.

Kluender, R., & Kutas, M. (1993). Subjacency as a processing phenomenon. *Language and Cognitive Processes, 8,* 573–633.

Kluender, R., & Münte, T. F. (1998). *Subject/object asymmetries: ERPs to grammatical and ungrammatical wh-questions*. 11th annual CUNY conference on human sentence processing, Rutgers University.

Kotz, S. A. (2008). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language, 109,* 68–74.

Kutas, M., & Iragui, V. (1998). The N400 in a semantic categorization task across 6 decades. *Electroencephalography and Clinical Neurophysiology. Evoked Potentials, 108,* 456–471.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.

Marinova-Todd, S. H. (2003). *Comprehensive analysis of ultimate attainment in adult second language acquisition*. Unpublished doctoral dissertation, Harvard University.

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language, 55,* 381–401.

Mueller, J. L., Hahne, A., Fujii, Y., & Friederici, A. D. (2005). Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. *Journal of Cognitive Neuroscience, 17,* 1229–1244.

Münte, T. F., Matzke, M., & Johannes, S. (1997). Brain activity associated with syntactic incongruencies in words and pseudo-words. *Journal of Cognitive Neuroscience, 9,* 318–329.

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience, 17,* 1212–1228.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9,* 97–113.

Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms*. Princeton: Princeton University Press.

Rodriguez-Fornells, A., Clahsen, H., Lleó, C., Zaake, W., & Münte, T. F. (2001). Event-related brain responses to morphological violations in Catalan. *Cognitive Brain Research, 11,* 47–58.

Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience, 18,* 2030–2048.

Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Sciences, 26,* 709.

Sabourin, L., & Stowe, L. A. (2004). Memory effects in syntactic ERP tasks. *Brain and Cognition, 55,* 392–395.

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research, 24,* 397–430.

Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research, 22,* 1–29.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research, 12,* 40–72.

Sebastián-Gallés, N., Martí, M., Carreiras, M., & Cuetos, F. (2000). *LEXESP: Léxicoinformatizado del español*. Barcelona: Ediciones Universitat de Barcelona.

Silva-Pereyra, J. F., & Carreiras, M. (2007). An ERP study of agreement features in Spanish. *Brain Research, 1185,* 201–211.

Singleton, D. (2005). The critical period hypothesis: A coat of many colours. *International Review of Applied Linguistics in Language Teaching, 43,* 269–285.

Tokowicz, N., & MacWhinney, B. (2005). Implicit vs. explicit measures of sensitivity to violations in L2 grammar: An event-related potential investigation. *Studies in Second Language Acquisition, 27,* 173–204.

Van Boxtel, S. (2005). *Can the late bird catch the worm? Ultimate attainment in L2 syntax*. Utrecht: Landelijke Onderzoekschool Taalwetenschap (LOT).

van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research, 22,* 241–255.

Vos, S. H., Gunter, T. C., Kolk, H. H., & Mulder, G. (2001). Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology, 38,* 41–63.

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience, 8,* 231–256.